

NBER WORKING PAPER SERIES

SPATIAL GROWTH AND INDUSTRY AGE

Klaus Desmet
Esteban Rossi-Hansberg

Working Paper 13302
<http://www.nber.org/papers/w13302>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2007

We thank Kunal Dasgupta for his excellent research assistance and Tom Holmes, Robert Lucas, Jordan Rappaport, Paul Romer, Giorgio Topa, and several seminar participants for useful comments. Desmet acknowledges the financial support of the Fundación Ramón Areces. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2007 by Klaus Desmet and Esteban Rossi-Hansberg. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Spatial Growth and Industry Age
Klaus Desmet and Esteban Rossi-Hansberg
NBER Working Paper No. 13302
August 2007
JEL No. E2,O3,O4,R1

ABSTRACT

U.S. county data for the last 20 or 30 years show that manufacturing employment has been deconcentrating. In contrast, the service sector exhibits concentration in counties with intermediate levels of employment. This paper presents a theory where local sectoral growth is driven by technological diffusion across space. The age of an industry -- measured as the time elapsed since the last major general purpose technology innovation in the sector -- determines the pattern of scale dependence in growth rates. Young industries exhibit non-monotone relationships between employment levels and growth rates, while old industries experience negative scale dependence in growth rates. The model then predicts that the relationship between county employment growth rates and county employment levels in manufacturing at the turn of the 20th century should be similar to the same relationship in services in the last 20 years. We provide evidence consistent with this prediction.

Klaus Desmet
Department of Economics
Universidad Carlos III de Madrid
C./ Madrid, 126
28903 Getafe (Madrid)
SPAIN
klaus.desmet@uc3m.es

Esteban Rossi-Hansberg
Princeton University
Department of Economics
Fisher Hall
Princeton, NJ 08544-1021
and NBER
erossi@princeton.edu

Spatial Growth and Industry Age*

Klaus Desmet
Universidad Carlos III and CEPR

Esteban Rossi-Hansberg
Princeton University

July 2007

Abstract

U.S. county data for the last 20 or 30 years show that manufacturing employment has been deconcentrating. In contrast, the service sector exhibits concentration in counties with intermediate levels of employment. This paper presents a theory where local sectoral growth is driven by technological diffusion across space. The age of an industry —measured as the time elapsed since the last major general purpose technology innovation in the sector— determines the pattern of scale dependence in growth rates. Young industries exhibit non-monotone relationships between employment levels and growth rates, while old industries experience negative scale dependence in growth rates. The model then predicts that the relationship between county employment growth rates and county employment levels in manufacturing at the turn of the 20th century should be similar to the same relationship in services in the last 20 years. We provide evidence consistent with this prediction.

1 Introduction

Improvements in technology to produce goods and services can come gradually, in incremental steps, or as large fundamental changes. When one of these fundamental changes occurs in a particular sector, producers in different regions need to adopt the new technology and adapt it to their particular environment and product. The result is a sequence of gradual changes in technology that improve productivity; first in areas that specialize heavily in this sector, then in other areas. Many of these gradual technological changes can be facilitated by the concentration of firms in the same region, as producers learn from each other how to implement the new general purpose technology (GPT). Eventually, producers in all regions change and adapt their production techniques, the technology in the industry matures, and diffusion stops. Then, the age of an industry —the time elapsed since the last new GPT in the industry— determines relative industry productivity of regions and the spatial distribution of employment across industries or sectors.

*We thank Kunal Dasgupta for his excellent research assistance and Tom Holmes, Robert Lucas, Jordan Rappaport, Paul Romer, Giorgio Topa, and several seminar participants for useful comments. Desmet acknowledges the financial support of the Fundación Ramón Areces.

We start this paper with one observation: between 1970 and 2000 the scale dependence of U.S. county employment growth exhibited very different patterns in manufacturing and services (see also Desmet and Fafchamps, 2006). County employment growth in manufacturing between 1970 and 2000 decreased with the level of employment in 1970. This negative correlation holds across the entire distribution. As can be seen in Figure 1, if we use a non-parametric estimate of this relationship, its slope is clearly negative for all county employment sizes. This implies that small counties, in terms of manufacturing employment, grew faster than large ones. Manufacturing employment thus dispersed in space. A different conclusion emerges when we look at services. Small counties, in terms of service employment, grew faster than larger ones, but the relationship reverses and becomes positive for intermediate counties, and then turns negative again at the high end of the distribution. As Figure 1 shows, we thus get an S-shaped pattern in services employment growth as a function of initial employment in services.

[FIGURE 1 ABOUT HERE]

What can account for this important difference across industries in the evolution of the distribution of economic activity? Our take is that industry age, and therefore the intensity of innovation, plays a role in explaining this disparity in the evolution of the two sectors. In a mature industry, the incentives to add employment in a particular location come mostly from the cost savings in land rents and other location specific factors. The benefits from agglomeration have been mostly exploited; similar to the mechanism at work in the “nursery cities” of Duranton and Puga (2001). As a result, manufacturing firms abandon regions with high densities of manufacturing employment in favor of other, more economical, regions. If land is a factor of production and goods are costly to transport, concentration of employment is costly. These costs are reflected in the wages firms have to pay at a particular location, as well as in land rents. Hence, in the absence of knowledge spillovers, production tends to spread evenly in space. As spillovers become less important, small counties will thus grow faster than large ones. This is consistent with the observed pattern of manufacturing employment growth between 1970 and 2000. The last main GPT in manufacturing — the introduction of electricity — dates back to the turn of the twentieth century. By the middle of the century this technology had been implemented everywhere and manufacturing became a mature industry.

In contrast, in the last thirty years, the recent revolution in information and communication technology has changed the way in which the service industry operates. This is a new GPT that has affected primarily the service sector and, using our definition, made it young. Many firms are still adapting to the new technological paradigm and much of its potential has yet to be exploited. Many lateral innovations are still to be made. Some regions, as the U.S. Midwest, that used to specialize

mostly in manufacturing, have started to switch towards service industries. Trade, of course, facilitates this phenomenon since it allows regions to specialize. Some counties transition from very low levels of employment in services to much higher levels, as the specialization of land use switches due to the benefits of trade. Counties with intermediate levels of service employment behave differently. Instead of production dispersing as land use patterns change, production agglomerates to take advantage of the still important benefits from spatial concentration. Hence, some intermediate counties, always in terms of service employment, grow faster than the smaller ones. The largest counties also obtain these benefits, but they also face the cost of congestion (or the decreasing benefits of agglomeration), which leads again to dispersion. The result is an S-shaped pattern as the one observed for several service industries for 1970-2000 in Figure 1.

In this paper we document these patterns in the data and present a theory that can rationalize them, using the arguments above. In this theory, industry age —as measured by the time passed since the last GPT innovation— is the key determinant of the shape of the relationship between the scale of employment in an industry at a location and its subsequent growth. Of course, this theory then implies that if we look at manufacturing employment growth at the turn of the twentieth century we should see a similar pattern to the one we observe in the service industry for the last 30 years. Figure 1 also presents the growth rates in manufacturing employment between 1900 and 1920 as a function of manufacturing employment in 1900.¹ As the theory predicts, the shape of manufacturing growth in 1900-1920 is almost identical to the shape in service employment growth between 1970-2000. Both sectors exhibit the same type of S-shaped pattern for these very different periods in time. This novel finding is what underlies and motivates our theoretical model.

In sum, at the turn of the twentieth century manufacturing in the U.S. experienced rapid innovation, prompted by the advent of electricity. Knowledge spillovers made geographic concentration of manufacturing employment useful. In contrast, by the end of the century manufacturing had matured and standardized. There were less benefits to be reaped from agglomeration. At the same time, however, knowledge spillovers gained in importance in the service industry, as improvements in information and communication technology caused a wave of product and process innovation in that sector. Our take is that the similarity between manufacturing at the turn of the twentieth century and services seventy years later is the result of innovation and spillovers being important for these young industries. In contrast, these effects had lost much of their importance for manufacturing at the end of the twentieth century.

¹Details of the calculation of these growth rates as well as confidence intervals are presented in Section 5. Since we detrend all data and growth rates are annual, all curves are comparable even though they include intervals of different length.

We present a two sector spatial theory that formalizes the logic we describe above. The theory inherits many characteristics present in Lucas and Rossi-Hansberg (2002) and Rossi-Hansberg (2005). The main difference is that technological change happens through a combination of diffusion and spillovers. In particular, we allow regions that do not benefit from knowledge spillovers to obtain the best technologies previously invented in more dense areas. In this sense, what leads to dynamics in the model will be the diffusion of technology from high density areas with large levels of knowledge spillovers and innovation to low density areas that do not obtain these benefits directly.

In order to qualitatively fit the main patterns described above, together with the rising share of employment in services, we allow locations to trade, and use constant elasticity of substitution (CES) preferences. Trade causes specialization. This is important in two respects. First, technological change can lead to changes in specialization patterns. By having some locations with initially no employment in an industry switch to that industry, we get fast growth in those locations, resulting in the first part of the S-shaped pattern. Second, specialization allows for different patterns of scale dependence in growth rates across sectors. Of course, as trade costs rise, these differences in scale dependence vanish. By adopting CES preferences, with an elasticity of substitution greater than one, this becomes less important, as the link between productivity growth and employment growth strengthens. In addition, these preferences allow us to fit the increasing service employment share in the U.S.

Our theory rationalizes the patterns in scale dependence of employment growth as the result of the changes in spillovers, diffusion and adoption of new technologies that result from general purpose innovations in a particular sector. It is the evolution of technology in a given sector what drives changes in sectoral employment. Hence, the theory implies that we should observe the same patterns of scale dependence when we look at productivity. In Section 5 we document that this seems to be the case in the data if we use the measure of TFP implied by our theory.

The empirical literature on technology diffusion goes back to the seminal work of Griliches (1957) who studied the spread of the use of hybrid corn in the U.S. In the specific case of diffusion of GPTs, not much attention has been given to the spatial dimension. Rosenberg and Trajtenberg (2004) describe how the replacement of waterwheels by steam engines allowed manufacturing activity to relocate from rural to urban areas. By removing the geographic constraint of proximity to water, firms could move to densely populated areas, where they could take advantage of agglomeration economies. Whether this clustering had anything to do with knowledge spillovers is unclear though. In the case of the Internet, Forman, Goldfarb and Greenstein (2005) find evidence pointing in that direction. While mere participation in the Internet became rapidly widespread across locations, the more complex applications, such as e-commerce, located predominantly in urban areas, where they had access to coinventions and complementary activities.

A further question of relevance raised by our work is the way in which IT is similar to electricity. Hobijn and Jovanovic (2001) and Jovanovic and Rousseau (2005) have pointed out the many similarities between both GPTs, ranging from their diffusion patterns to the behavior of IPOs, patents and the stock market. In the midst of disappointment about the computer revolution, David (1990) used the experience of the electric dynamo to argue there was nothing surprising about the productivity slowdown paradox. In a recent contribution Atkeson and Kehoe (2006) are more cautious. Based on a calibrated model of the electricity revolution, they find that slow diffusion depends crucially on agents having built up a large stock of knowledge about the old technology. For lack of data on the IT era, they conclude that it remains to be seen whether the computer will be a simple replay of the dynamo. Our paper suggests that at least along the dimension of spatial growth, electricity and IT exhibit similar behavior.

Although not its main focus, our model is consistent with the rise of the service economy. The literature has explained this structural transformation by relying on either non-homothetic preferences (Kongsamut, Rebelo and Xie, 2001) or uneven technological progress (Ngai and Pissarides, 2004). Regarding this latter view, homotheticity in preferences can be maintained as long as the elasticity of substitution differs from unity. Ngai and Pissarides (2004) assume an elasticity of substitution below one, implying that TFP growth in manufacturing must be faster than in services to account for the structural transformation. Our model assumes an elasticity of substitution above one, so that the structural transformation is consistent with higher TFP growth in services.² Although Baumol (1967) argued that services were bound to inherently trail behind manufacturing in terms of productivity growth, there is some debate about whether his premise continues to hold. Bosworth and Triplett (2006) provide evidence that by the second half of the 1990s TFP growth in services outstripped that in manufacturing, and accordingly declared ‘Baumol’s Disease’ cured. Independently of this debate, the essential driving force underlying our conclusions for the scale dependence patterns in young and mature sectors is *not* the absolute level of TFP growth, but rather the cross-county variation in TFP growth. It is this regional variation in growth rates which is the main focus of our analysis.

There has been little work in dynamic models of spatial industry location. Rossi-Hansberg (2004) introduces a similar framework with capital accumulation. Holmes (2004) provides a dynamic model where cluster location changes across time as firms take advantage of location specific factors. There is a variety of frameworks that study the distribution of city sizes using dynamic models (Gabaix, 1999; Duranton, 2006; Rossi-Hansberg and Wright, 2006), or focusing on rural to urban migration in a dynamic setting (Lucas, 2004; Henderson 2005), but none of these papers uses a spatial theory and so they have no prediction for sectoral employment growth across regions. Henderson and Venables

²The same implications about productivity, in a very different model, are present in Buera and Kaboski (2006).

(2005) present a dynamic model of city evolution, but likewise without a spatial component.

The rest of the paper is organized as follows. In Section 2 we present the model. Section 3 presents numerical result that characterize an equilibrium for the case with trade. Section 4 presents the evidence and Section 5 concludes.

2 The Model

We model a spatial economy in a closed interval $[0, 1]$. The density of land at each location ℓ is one. Agents own land where they work and live and receive the corresponding rents. Apart from land, there is no other saving technology.

2.1 Preferences

Agents live where they work and they derive utility out of consumption of two goods: manufactures and services. Each location can produce in both sectors or specialize in one of them. Labor is freely mobile across locations and sectors, so that the indirect utility that agents derive at each location has to be identical. Mobility across sectors further implies that wages in all sectors will be the same in a given location. All agents are endowed with one unit of labor which they supply inelastically. Agents order consumption bundles according to a utility function $U(c_M, c_S)$ with standard properties. We also assume that $U(\cdot)$ is homogenous of degree one. c_i denotes consumption of good $i \in \{M, S\}$.

The problem of a consumer at a particular location ℓ is given by

$$\begin{aligned} & \max_{c_i} U(c_M, c_S) \\ \text{s.t.} \quad & w(\ell) + R(\ell)/L(\ell) = p_M(\ell) c_M(\ell) + p_S(\ell) c_S(\ell) \end{aligned}$$

where $p_i(\ell)$ denotes the price of good i at location ℓ , $R(\ell)$ denotes land rents at location ℓ (so $R(\ell)/L(\ell)$ is the dividend from land ownership since $L(\ell)$ denotes total employment at ℓ and all agents are identical), and $w(\ell)$ the wage at ℓ . The first order conditions of this problem yield $U_i(c_M, c_S) = \lambda(\ell) p_i(\ell)$, for all $i \in \{M, S\}$, where $U_i(\cdot)$ denotes the marginal utility of consuming good i and $\lambda(\ell)$ is a location-specific Lagrange multiplier. Denote by $\bar{U}(p_M(\ell), p_S(\ell), w(\ell) + R(\ell)/L(\ell))$ the indirect utility function of an agent at location ℓ . Because of free mobility of labor, it must be the case that

$$\bar{U}(p_M(\ell), p_S(\ell), w(\ell) + R(\ell)/L(\ell)) = \bar{u}, \text{ for all } \ell \in [0, 1],$$

where \bar{u} is determined in equilibrium.

We will use the CES utility function to illustrate the results of the model, so we let

$$U(c_M, c_S) = (h_M c_M^\beta + h_S c_S^\beta)^{1/\beta}$$

with $\beta < 1$. Thus $h_i c_i^{\beta-1} \bar{u}^{1-\beta} = \lambda(\ell) p_i(\ell)$, for all $i \in \{M, S\}$, where $\lambda(\ell) = \frac{\bar{u}}{w(\ell) + R(\ell)/L(\ell)}$ and

$$c_i = \left(\frac{h_i (w(\ell) + R(\ell)/L(\ell))}{p_i(\ell) \bar{u}^\beta} \right)^{\frac{1}{1-\beta}} \quad \text{all } i \in \{M, S\}.$$

Hence, wages and prices need to satisfy

$$w(\ell) + R(\ell)/L(\ell) = \bar{u} \left(h_M \left(\frac{h_M}{p_M(\ell)} \right)^{\frac{\beta}{1-\beta}} + h_S \left(\frac{h_S}{p_S(\ell)} \right)^{\frac{\beta}{1-\beta}} \right)^{-\frac{1-\beta}{\beta}}.$$

2.2 Technology

The manufacturing sector is assumed to be more land intensive than the service sector and both sectors face knowledge spillovers. The inputs of production are land and labor. Production per unit of land in the manufacturing sector is given by

$$M(L_M(\ell)) = Z_M(\ell) L_M(\ell)^\mu,$$

and, similarly, in the service sector we have

$$S(L_S(\ell)) = Z_S(\ell) L_S(\ell)^\sigma,$$

where $Z_i(\ell)$ denotes total factor productivity in sector i and location ℓ and $L_i(\ell)$ is the amount of labor per unit of land used at location ℓ in sector i . We assume that $\mu < \sigma < 1$, so the manufacturing sector is more land intensive than the service sector. As we will specify below, total factor productivity (TFP) in each sector depends on the amount of labor employed in the same sector in neighboring locations. We assume that a firm takes $Z_i(\ell)$ as given and so the effect of other producers on productivity is not taken into account by the firm: an externality. Thus the problem of a firm in sector $i \in \{M, S\}$ at location ℓ is given by

$$\max p_i(\ell) Z_i(\ell) L_i(\ell)^\iota - w(\ell) L_i(\ell),$$

where $\iota \in \{\mu, \sigma\}$. The first order conditions yields $\iota Z_i(\ell) \hat{L}_i(\ell)^{\iota-1} = \frac{\iota M(\hat{L}_i(\ell))}{\hat{L}_i(\ell)} = \frac{w(\ell)}{p_i(\ell)}$, which implies that

$$\hat{L}_i(\ell) = \left(\frac{\iota Z_i(\ell) p_i(\ell)}{w(\ell)} \right)^{\frac{1}{1-\iota}},$$

or

$$w(\ell) = \frac{\iota Z_i(\ell) p_i(\ell)}{\hat{L}_i(\ell)^{1-\iota}}.$$

The bid rent in sector i is given by $R_i(\ell) = p_i(\ell) Z_i(\ell) \hat{L}_i(\ell)^\iota - w(\ell) \hat{L}_i(\ell)$ and so

$$R_i(\ell) = \left(\frac{1-\iota}{\iota} \right) \hat{L}_i(\ell) w(\ell) = (1-\iota) \left(\frac{\iota}{w(\ell)} \right)^{\frac{\iota}{1-\iota}} (Z_i(\ell) p_i(\ell))^{\frac{1}{1-\iota}}.$$

We still need to specify how TFP is determined in each industry. We let

$$Z_i(\ell) = \max \left[\begin{array}{c} \rho \bar{Z}_i^{\max} + (1 - \rho) \left(\int_0^1 e^{-\delta_i |\ell - r|} \hat{L}_i(r) \theta_i(r) dr \right)^{\gamma_i} \\ \left(\int_0^1 e^{-\delta_i |\ell - r|} \hat{L}_i(r) \theta_i(r) dr \right)^{\gamma_i} \end{array} \right],$$

where $\theta_i(\ell)$ denotes the fraction of land at location ℓ used in the production of good i and $\rho \in [0, 1]$. If $\rho = 1$ all locations have access to the general level of technology \bar{Z}_i^{\max} . (If $\rho < 1$ the general level of technology also diffuses, but not perfectly.) Locations can possibly improve upon this general level of technology, \bar{Z}_i^{\max} , if they benefit from sufficiently large spillovers. These spillovers are a weighted average of employment at all locations where the weights are a function of distance. The best technology becomes public domain in the next period. In other words, the general level of technology in a given period is the maximum level of technology across all locations in the previous period. We will describe these dynamics in further detail later on. We assume that $\gamma_i + \max[\mu, \sigma] < 1$. This guarantees that spillovers are a concave function of total population and so economic activity does not agglomerate in only one point. It also implies that very dense locations gain less from extra workers than less dense locations: a form of congestion.

2.3 Static equilibrium in the absence of trade

In equilibrium labor markets clear. Given free mobility, we have to guarantee that the total amount of labor in the economy is equal to the total supply \bar{L} . The labor market equilibrium condition is therefore $\int_0^1 \sum_i \theta_i(r) \hat{L}_i(r) dr = \bar{L}$.

In the absence of trade all locations have to produce both goods since the marginal utility of consuming the first unit of any good is infinity. This implies that at all locations ℓ the returns to land must equalize across sectors, so that $R(\ell) \equiv R_M(\ell) = R_S(\ell)$. Hence,

$$\left(\frac{1 - \sigma}{\sigma} \right) \hat{L}_S(\ell) = \left(\frac{1 - \mu}{\mu} \right) \hat{L}_M(\ell)$$

and by definition $L(\ell) = \sum_i \theta_i(\ell) \hat{L}_i(\ell)$. This implies that

$$\hat{L}_i(\ell) = \frac{\frac{\ell}{1-\ell}}{\hat{\theta}(\ell)} L(\ell),$$

where

$$\sum_i \theta_i(\ell) \frac{\ell}{1-\ell} = \hat{\theta}(\ell).$$

Land rents and prices are then given by

$$R(\ell) = \frac{L(\ell) w(\ell)}{\hat{\theta}(\ell)}$$

and

$$p_i(\ell) = \frac{w(\ell)}{\iota Z_i(\ell) \left(\hat{L}_i(\ell) \right)^{\iota-1}}.$$

Using the above expression for land rents, we know that $w(\ell) + R(\ell)/L(\ell) = \left(1 + 1/\hat{\theta}(\ell)\right) w(\ell)$, and so utility levels are given by

$$\bar{u} = \left(1 + \hat{\theta}(\ell)\right) \left[\sum_i h_i^{\frac{1}{1-\beta}} \left(\frac{\iota Z_i(\ell) \left(\frac{\iota}{1-\iota} L(\ell) \right)^{\iota-1}}{\hat{\theta}(\ell)^{-\iota}} \right)^{\frac{\beta}{1-\beta}} \right]^{\frac{1-\beta}{\beta}}.$$

Production in location ℓ in industry i is $Z_i(\ell) (L_i(\ell))^\iota$ and the product market equilibrium conditions are $\theta_i(\ell) Z_i(\ell) L_i(\ell)^\iota = c_i(\ell) L(\ell)$. Hence,

$$\theta_i(\ell) = \left(\frac{h_i \iota \left(\frac{\iota}{1-\iota} \right)^{\iota\beta-1}}{\bar{u}^\beta} \right)^{\frac{1}{1-\beta}} Z_i(\ell)^{\frac{\beta}{1-\beta}} \left(\left(1 + \hat{\theta}(\ell)\right) \hat{\theta}(\ell)^{-\iota\beta} \right)^{\frac{1}{1-\beta}} L(\ell)^{\frac{-\beta(1-\iota)}{1-\beta}}$$

where $\hat{\theta}(\ell)$ is determined by $\sum_i \theta_i(\ell) = 1$.

The θ_i expression implies that the fraction of land used for production in each industry, and therefore the share of labor used in that industry, depends on the productivity in the sector. Once we introduce dynamics, productivities may grow at different rates across sectors, in which case the scale dependence of growth rates will differ across sectors. This will be key for explaining the different growth patterns in manufacturing and services. If we had chosen Cobb-Douglas preferences instead, the shares of labor and land allocated to the different sectors would have been constant across locations and time. Therefore, while CES is able to generate differences in scale dependence across sectors, Cobb-Douglas is not.

2.4 Transport costs and trade

We now introduce trade. This will allow regions to specialize as in Rossi-Hansberg (2005). Goods are costly to transport. For simplicity we assume iceberg transportation costs that are identical in all industries (the latter assumption is without loss of generality given that the equilibrium only depends on the sum of transport costs in both industries).³ If one unit of any of the goods is transported from ℓ to r only $e^{-\kappa|\ell-r|}$ units of the good arrive in r . Note that since the technology to transport goods is freely available this implies that if good i is produced in location ℓ and consumed in location r the price of the good has to satisfy $p_i(r) = e^{\kappa|\ell-r|} p_i(\ell)$.

³To see this, suppose transportation costs are zero in services and prohibitive in manufacturing. Because we impose trade balance location by location, services would become equally nontradeable as manufactured goods, in spite of being freely transportable. The difference in transportation costs across sectors therefore plays no role in the current model.

Land is assigned to its highest value. Hence, land rents are such that $R(\ell) = \max\{R_M(\ell), R_S(\ell)\}$. Therefore, in general, land will specialize in one of the two potential uses. So if $R(\ell) = R_i(\ell)$, then $\theta_i(\ell) > 0$. Of course, with complete specialization this condition becomes $\theta_i(\ell) = 1$.

In order to guarantee equilibrium in product markets when there is trade, we need to take into account that some of the goods are lost in transportation. To do this define the stock of excess supply, $H_i(r)$, of product i between locations 0 and r by

$$H_i(0) = 0$$

and the differential equation

$$\frac{\partial H_i(r)}{\partial r} = \theta_i(r) x_i(r) - \hat{c}_i(r) \left(\sum_i \theta_i(r) \hat{L}_i(r) \right) - \kappa |H_i(r)|.$$

That is, at each location we add to the stock of excess supply the amount of good i produced (where $x_M(r) = M(\hat{L}_M(r))$ and $x_S(r) = S(\hat{L}_S(r))$ denote equilibrium production of good i at location r per unit of land) and we subtract the consumption of good i (by all residents of r). We then need to adjust for the fact that if $H_i(r)$ is positive, as we increase r we have to ship the stock of excess supply a longer distance which implies a cost in terms of goods and services that is given by κ . The equilibrium conditions in the goods markets are then given by

$$H_i(1) = 0 \text{ for all } i.$$

We will impose trade balance location by location. This implies that the value of the goods that are shipped to location ℓ must be identical to the value of the goods that are shipped from location ℓ , so that $p_M H_M(\ell) + p_S H_S(\ell) = 0$. The trade balance condition says that the value of goods produced and consumed at ℓ is equal, once transport costs in terms of goods are covered.

Lucas and Rossi-Hansberg (2002) and Rossi-Hansberg (2005) show that a static equilibrium of this economy exists for an arbitrary level of technology in both sectors. Hence, since the dynamics only operate through the level of technology in the different regions, an equilibrium of this model exists as well.

2.5 Evolution of Technology

So far we have ignored the time dimension completely (including in the notation). The dynamics of the model will come only through changes in the general level of technology and so an equilibrium of the dynamic economy is just a series of static equilibria described above but with different distributions of technology across periods. We index all variables by time and suppose the economy has a general

level of technology at time t given by $\bar{Z}_i^{\max}(t)$ for all i . Then, next period's level of technology is given by

$$\bar{Z}_i^{\max}(t+1) = \max_{\ell} Z_i(\ell, t).$$

Assume $\rho = 1$. In this case, after one period the better technology that some producers obtain from the benefits of agglomeration is public domain. That is, all producers in all locations have access to the best technology of the previous period.

Given the specification of $Z_i(\ell, t)$ and $\rho = 1$, this implies that $\bar{Z}_i^{\max}(t+1)$ is a weakly increasing sequence. It also implies that technology can be in steady state in some industries in which there is no innovation. If

$$\bar{Z}_i^{\max}(t+1) = \bar{Z}_i^{\max}(t) \text{ all } i,$$

the economy has converged to the steady state. Since the model only exhibits transitional dynamics, growth rates of employment in all sectors in all locations become zero. Note, however, that if $\bar{Z}_i^{\max}(t+1) = \bar{Z}_i^{\max}(t)$ in only one industry, there may be non-zero growth rates in all industries. Note that technology in both sectors is bounded by \bar{L}^γ , that is, $\lim_{t \rightarrow \infty} \bar{Z}_i^{\max}(t) \leq \bar{L}^\gamma$, all i . Since for all i , $\lim_{t \rightarrow \infty} \bar{Z}_i^{\max}(t)$ is a weakly increasing sequence, these sequences have to converge.

We think of GPTs as a new general production technology that induces technological innovation, spillovers and diffusion in an industry, following a period in which technological innovation had essentially come to a halt. In the wake of a GPT, technology is determined by the process described above. As technology in an industry matures, spillovers are no longer useful and the technology in that industry reaches a steady state. Then a new GPT may emerge benefitting another industry. Thus, in the beginning of the twentieth century manufacturing was the young sector and services the old one. In this context it is natural to think of the younger sector having the higher level of technology. So we start the economy assuming that $\bar{Z}_S^{\max} < \bar{Z}_M^{\max}$. This reverses by the end of the century, as manufacturing is now old and services young. Of course, if we want to calibrate the economy we need other parameters that determine the units of each of these products. Since we are interested in growth rates, in order to save on notation we do not add these scale factors.

3 Numerical Results

The following numerical simulations of the model use the parameters in Table 1. All parameters have been defined above except the last four. The parameters in the model that are not in Table 1 are kept equal to one. We need to start the economy with an initial level of technology. Instead of assuming a constant level, in both sectors we use an initial productivity which is quadratic and symmetric around $\ell = 0.5$, where it obtains its maximum, $\max_{\ell} Z_i(0)$. We use a quadratic instead of a constant function

to simplify the computations and avoid multiplicity of equilibria (e.g. manufacturing areas can be at the boundaries or at the center).

Table 1: Parameter Values

Parameter	Value	Parameter	Value
β	0.5	ρ	0.95
μ	0.5	$\min Z_M(0)$	0.39
σ	0.55	$\max Z_M(0)$	0.40
γ_i	0.05	$\min Z_S(0)$	0.19
δ_i	10	$\max Z_S(0)$	0.20
κ	0.005		

The resulting equilibrium has three regions of specialization and it is symmetric. The areas at the boundaries of the interval $[0, 1]$ specialize in manufacturing and the region in the middle in services. The size of these regions changes as technology evolves, but this pattern is constant across periods. As shown in Rossi-Hansberg (2005) higher trade costs would result in more switches in land specialization.

For computational simplicity we fix $\bar{u} = 1$ and let total population size adjust, rather than fixing the population size and adjusting the utility levels. Doing this eliminates one of the iterations needed to compute an equilibrium and makes the computations feasible. Thus, one should interpret our economy as one with migration and population growth. We also distribute rents to absentee landlords instead of rebating them to consumers in the region.

We want to illustrate the implications of the model for sectoral growth rates and for the relative shares of manufacturing and services. We have chosen parameters that will allow us to do this in a simple way. Since the main general purpose innovation in manufacturing occurred at the beginning of the twentieth century, much earlier than in services, we set initial productivity in manufacturing to be higher than in services. Starting from those values, in an initial stage we let productivity in manufacturing evolve as implied by the model, whereas we keep productivity in the service sector fixed. This reflects the fact that manufacturing is young and firms are adapting to the general purpose technology. Then, at a later stage, we introduce a new GPT that impacts the service sector. From that point onwards we let the productivity in both the manufacturing and the service sector evolve as described in Section 2.

We present results for employment growth in both industries for three distinct time periods: before the GPT in services; just after the GPT in services; and some time later when technologies in both sectors are mature. The first period should illustrate one in which the GPT in manufacturing happened recently, but there is not much innovation in services (as in 1900-1920 in the data). The second period is one in which manufacturing is already a mature industry, and innovation is very active in the service industry (as in 1980-2000 in the data). The third period illustrates the case when

both technologies are mature (the future, absent new GPTs).

Figure 2 presents the results for manufacturing and service growth in these three time periods. (We present only three periods but we iterate several times between periods.) The solid curve shows the growth rate in manufacturing employment ordered by the size of manufacturing employment in that location. The dashed curve shows the same information for the service sector. The length of a given interval on the horizontal axis measures a set of locations specializing in a given industry.⁴ Due to trade all locations specialize, so the set of locations for which employment is positive differs across sectors. Of course, the sum of the lengths of the intervals for manufacturing and services adds up to one. In all cases we present a trend calculated by fitting a polynomial of degree six. The polynomial of degree six allows us to smooth employment growth rates in a sector and present just the scale dependence implied by the model, without the noise that results from having several distinct growth rates associated with a particular employment size. This issue is present only in the manufacturing industry in Periods 2 and 3, since this industry is losing employment in these periods in locations with the same level of employment as locations that keep producing manufacturing goods. In all other graphs this issue is not present but we apply the same smoothing for comparison purposes. This smoothing also eliminates some flat parts at the minimum (or the maximum) of these relationship which indicate growth rates equal to infinity or minus one. These are locations at which land use switches sector. All growth rates in Figure 2 are relative to the mean growth rate in the period, so we present growth rates after subtracting the average growth rate in the sector and period.

It is clear in Figure 2a for period 1 that the model produces the non-monotonicities (S-shaped) pattern observed in the data for manufacturing in 1900-1920. Locations with no or little manufacturing employment exhibit convergence, with smaller locations experiencing faster growth. These are locations, specialized in services and isolated from existing manufacturing clusters, that switch to manufacturing production. Since they do not benefit from spillovers from neighboring locations, the standard convergence forces are at work. Locations with already some manufacturing employment exhibit divergence, with smaller locations growing slower than larger ones. Because of agglomeration forces, the larger ones tend to have larger neighbors, and benefit from greater spillovers, the driving force for the concentration of employment. Once locations become large enough, the pattern switches around once again, and convergence is back: Congestion dominates spillovers from neighboring locations, and employment disperses.

⁴The horizontal axis in Figure 2 exhibits the location index ordered by employment size, not employment size itself. This allows us to represent in the figure also the length of the intervals that specialize in each industry. In all industries and periods, the graphs using actual employment size in the industry exhibit the same patterns of scale dependence.

According to our simulations, during period 1 employment growth in the service sector is essentially zero (although actual growth rates are not exactly zero). Of course, this is the result of our assumption that technology in the service sector is fixed in the first period. Unfortunately there is no data of employment growth in services at the county level at the beginning of the twentieth century that is able to discipline our theoretical exercise for services in the first period. The service sector loses some locations to the manufacturing sector though. This is natural, as manufacturing technology is improving and diffusing.

[FIGURES 2a, 2b and 2c ABOUT HERE]

Just before the second period, the technology in the service industry starts evolving as described in Section 2. This reflects the impact of a new GPT. By now manufacturing has become a mature industry, although its technology is still diffusing. Its employment is becoming more disperse in order to take advantage of low land rents. In contrast, the service industry now exhibits the S-shaped pattern or non-monotonicity previously observed in the manufacturing sector. This seems to be the period that approximates best the data for the US between 1970-2000.

In the third period, both industries are mature, technology is relatively uniform, although diffusion continues, particularly in the now old (but relatively younger) service sector. Both industries are still taking advantage of low land rents in certain areas that used to have inferior technologies. Employment is becoming more dispersed. Growth rates in both industries decline with the size of employment in the corresponding sector. If we let technology diffuse perfectly, eventually growth rates become zero and the economy stops growing. This cycle is almost complete in the manufacturing sector, as can be observed in Figure 2c.

The focus in Figure 2 is on the shape, and not the level, of the curves. We de-trend all growth rates as we do in the data. The reason is that we can arbitrarily change the level by changing the initial technologies $\min Z_i(0)$ and the utility level \bar{u} . However, changing these parameters proportionally does not change the qualitative features of these curves. Note that the difference in growth rates across regions seem smaller in the model than in the data (see Section 5). Of course, this depends on the definition of a period. If we group several periods we can obtain differences of the same magnitude.

Output growth follows very similar patterns given that employment growth is solely the product of productivity growth in a particular sector and location. So it is interesting to look at the evolution of productivity at these three points in time. The following two figures show the level of productivity in manufacturing and services across the interval $[0,1]$. Since both sectors exhibit substantial productivity growth between the first and the second period, we use different axes for the first period (left axis) and for the second and third periods (right axis).

[FIGURE 3 ABOUT HERE]

Figure 3 shows how technology diffuses in the model. In the first period the productivity distribution in manufacturing is mostly the outcome of spillovers.⁵ In the second period, diffusion is the dominating force, except for the few locations exhibiting a bump. In those areas spillovers continue to prevail because of the high concentration of manufacturing employment in the surroundings. In particular, the productivity that results from agglomeration effects is higher than the productivity these regions had access to through technological diffusion. By the third period this more advanced technology has smoothly diffused to the other regions, as everyone has access (up to ρ) to the best technology used last period.

Figure 4 shows the evolution of TFP in the service sector. In this sector diffusion is strong enough so that technology increases substantially in all locations. By the third period technology is relatively uniform across locations. Note the differences in the scale of the axes, services is a younger industry and in the third period the differences in technology are much larger than in manufacturing.

[FIGURE 4 ABOUT HERE]

One dimension that the CES structure of the model can help us understand is the evolution of the shares of employment in the U.S. Because technology changes over time the relative price of manufactures and services changes as well. CES preferences then imply that the average (across locations) shares of consumption will change accordingly. This effect interacts with our parameter choices which make the service sector relatively more labor intensive. The result is a share of employment in services that is larger than in manufacturing and increases as the service sector starts innovating and these innovations diffuse in space. In contrast, in the first period the manufacturing employment share increases since manufacturing in that period is the young innovative sector. Figure 5 presents the shares of employment in both sectors. Key for these results is that $\beta = 0.5$ and so the elasticity of substitution in consumption between services and manufactures is larger than one. This figure should be compared with Figure 10 in Section 5 where we present the evolution of the same shares in the U.S. for the twentieth century.⁶

[FIGURE 5 ABOUT HERE]

⁵Note that because of specialization not all locations actually have manufacturing employment. In particular, the region in the middle is specialized in services. In that region the manufacturing productivity represents what the productivity would be were the region to produce manufactured goods.

⁶For other explanations of the increase in the share of employment in services in the second half of the 20th century see for example Buera and Kaboski (2006).

The results above were computed for a particular set of parameter values. The general pattern of scale dependence is, however, robust to many of these parameters and, in general, depends more on the relative values we choose in manufacturing versus services rather than on their level. A couple of parameters that may seem relatively low are the employment shares in manufacturing and services, μ and σ . If we change σ to 0.5 or 0.6 we obtain the same qualitative results. In particular we obtain the S-shape pattern in the second period. Figure 6 presents results for the growth rate in service employment in period 2. Clearly, the S-shape pattern becomes more pronounced the larger the value of σ . The results for different manufacturing employment shares, μ , behave similarly. The larger μ , the more pronounced the S-shape pattern observed in manufacturing employment growth in the first period.

[FIGURE 6 ABOUT HERE]

One potential concern with the results above is the initial productivity function we have chosen in both sectors. As mentioned, we start both sectors with a quadratic productivity function. The shape of these curves does affect the results. However, as we make the curvature smaller, so the initial variation in productivities across regions decreases, we obtain the same pattern of scale dependence but with amplified S-shaped patterns.

4 Empirical Evidence

Our theory predicts that spatial employment growth follows different patterns for young and mature industries. In young industries spatial employment growth is non-monotonic: locations with low or high employment exhibit convergence (deconcentration), whereas locations with intermediate levels of employment exhibit divergence (concentration). In mature industries spatial employment growth becomes monotonic: all locations, independently of their size, exhibit convergence (deconcentration). In this section we provide empirical evidence that supports these theoretical predictions and complements Figure 1 in the introduction. To do so, we compare spatial employment growth in young and mature sectors.

4.1 Industry Age

According to our theory, an industry is young when it benefits from the diffusion of a new technology and knowledge spillovers are strong. In contrast, an industry is mature when the technology it uses has become standardized and knowledge spillovers have lost much of their importance. By a new technology we do not mean a marginal improvement over an already existing technology, but rather a

radical innovation that represents a drastic change with the previously used technology. As mentioned before, this is akin to the introduction of a new General Purpose Technology (GPT).

To determine whether a given industry is young at a given point in time, we therefore refer to the large literature on GPTs. As argued by Jovanovic and Rousseau (2005) and David and Wright (2003), the two major GPTs in the 20th century were electricity and information technology (IT). Jovanovic et al. (2005) define the starting point of a GPT as the date at which it reaches 1% diffusion in the median sector. This gives a starting point of 1894 for electricity, coinciding with the first hydroelectric facility at Niagara Falls, and 1971 for IT, coinciding with Intel's 4004 microprocessor. Hobijn and Jovanovic (2001) provide additional evidence supporting the early 1970s as the starting point for IT. They argue that the decline in the stock market at the beginning of the 1970s coincided with the arrival of 'good news' about IT. Stock prices declined most in those sectors that had the largest post-1973 investments in IT, the idea being that in those sectors part of the capital stock became obsolete. As definition for the ending point of a GPT, Jovanovic et al. (2005) use the date at which the diffusion curve flattens. This gives an ending date of 1929 for electricity, whereas for IT the curve has not plateaued yet. This timing is consistent with many other events associated with GPTs. For example, in the 20th century there were two surges in patents and trademarks, the first one between 1900 and 1930 and the second one after 1977. Similarly, IPOs increased between 1895 and 1929, and again after 1977.

This evidence suggests that the diffusion of electricity started in earnest somewhere between 1894 and 1900, and ended by 1930, whereas the diffusion of IT started some time between 1971 and 1977, with the end is still not in sight. Given that the diffusion of IT has not yet plateaued, we cannot compare the ending date of both GPTs. To make the timing of both technologies comparable we focus on 1900-1920 and 1980-2000 as the periods when electricity and IT were young.

Although GPTs are pervasive in the sense that they tend to spread to the entire economy,⁷ their effect may differ depending on the sector. In the case of electricity, David and Wright (2003) argue that it affected mainly the manufacturing sector. In the decade after World War I, Kendrick (1961) estimates that economy-wide TFP grew by 22 per cent, whereas in manufacturing TFP grew by 76 per cent. In the case of IT, the evidence points to the service sector being the big beneficiary. Hobijn and Jovanovic (2001) compute IT intensity —the share of IT equipment in the total stock of equipment— in different sectors. In 1996 IT intensity stood at 42.4% in services, compared to 17.9% in manufacturing.⁸ Within the broad category of services, the subsectors that have invested most in IT

⁷Pervasiveness is one of the fundamental characteristics of GPTs, according to the definition of Bresnahan and Trajtenberg (1995).

⁸Similar results are found by Chun et al. (2005), Triplett and Bosworth (2002) and Basu and Fernald (2006).

are: FIRE (finance, insurance and real estate), communications, business services, and wholesale. In a growth accounting exercise of 60 industries, Bartelsman and Beaulieu (2004) find that the contribution of IT to growth has been most prevalent in credit institutions. Basu and Fernald (2006), for their part, suggest that the most IT intensive sectors are communications, finance and insurance, and business services. Chun et al. (2005) adds wholesale to that list. Not all of these sectors were equally fast to adopt IT. According to Hobijn and Jovanovic (2005) and Bartelsman and Beaulieu (2004), the early investors in IT were FIRE and communications.

Based on the timing and the differential sectoral effects, our empirical implementation takes the following stylized view. The manufacturing sector was young in the period 1900-1920, and mature after the 1950s, whereas the service sector was mature before 1970, and became ‘young’ some time in the period between 1970 and 1980. We therefore take both a time series approach (by following industries over time) and a cross-sectional approach (by comparing different industries) in studying spatial growth patterns. To guarantee that our evidence is not specific to the United States, we also analyze European data.

4.2 Manufacturing and Services in the United States

Our empirical analysis for the United States takes counties as the unit of observation. There are several reasons for doing so. First, given our focus on spatial growth, counties provide an appropriate level of geographic detail. Second, counties cover the entire U.S., in contrast to, for instance, metropolitan areas or cities. Third, county data allow us to go sufficiently back in time. Our data on county-level employment come from a variety of sources. Data until 1930 come from the Historical Census Browser at the University of Virginia; and data from 1969 onwards come from the Regional Economic Information System (REIS) compiled by the U.S. Bureau of Economic Analysis (BEA). One obvious concern are changing definitions of counties and county borders. By using information on years in which county definitions changed,⁹ our regressions exclusively focus on counties whose definitions have not changed over the period of interest.¹⁰

Given our focus on possible non-monotonicities in spatial employment growth, we run nonlinear kernel regressions of the form:

$$L_{t+s}^i = \phi(L_t^i) + e_t^i$$

where L_t^i is log employment in year t and county i . The estimation uses an Epanechnikov kernel with

⁹Information about changes in county borders come from Forstall (1996).

¹⁰In particular, depending on the regression, we leave out counties of which borders changed after 1900 or after 1980. To make sure that the different patterns between 1900-1920 and 1980-2000 are not due to different samples of counties, we re-ran all our regressions on the sample of counties of which borders did not change after 1900. None of the results changed.

optimal bandwidth.¹¹ Because the distribution of employment levels is approximately log-normal, we focus on the log of employment.

To facilitate interpretation, we plot annual employment *growth* as a function of initial log employment in the same industry. In this case, a negative slope indicates deconcentration (convergence) and a positive slope indicates concentration (divergence).

[FIGURE 7 ABOUT HERE]

Figure 7 plots de-trended annual employment growth in the manufacturing sector for the periods 1900-1920 and 1980-2000 together with the 95% confidence intervals.¹² As can be seen, for the period 1900-1920, when electricity was diffusing and manufacturing was young, employment growth was non-monotone. At that time there was deconcentration or convergence in the lower and the upper tail of the distribution, and concentration or divergence in the middle part of the distribution. In contrast, for the period 1980-2000, when manufacturing was mature, there was deconcentration or convergence across the entire distribution. Though not reported in the paper, when analyzing the years between 1920 and 1980, one can observe how the S-shaped curve gradually changes into a downward sloping curve.

A similar picture to the one in manufacturing at the beginning of the twentieth century emerges when analyzing the recent experience of the service sector. During the last two decades of the 20th century service industries invested heavily in IT, so we define the service sector as young for this period. Figure 8 shows de-trended growth rates of employment in service industries from 1980 to 2000 as a function of employment in the county in 1980. The figure also shows 95% confidence intervals. Comparing the growth rate in services in 1980-2000 to the same curve in manufacturing in 1900-1920 one can observe that the main difference is that small counties in manufacturing grew very fast. The reason is that there are very small counties, in terms of manufacturing employment, in 1900. In contrast, in 1980 almost all counties have a basic employment level in services of about 50 employees. Apart from this, both figures exhibit the exact same pattern of scale dependence.

If the IT revolution started somewhere in the middle of the 1970s, making services young, then during the decades before the 1970s services should have been mature, and thus exhibited negative scale dependence across the distribution. Although we do not have comparable data for that period, we checked for both retail and FIRE for the period 1950-1970, and found this prediction to hold up in the data.

¹¹This methodology is described in detail in Desmet and Fafchamps (2006).

¹²Figure 1 shows results starting in 1970, not 1980, which illustrates that these facts are robust to the precise choice of the latter period starting date.

[FIGURE 8 ABOUT HERE]

One potential concern is that the patterns we describe are mostly about metropolitan counties. Maybe what we are witnessing is simply industries moving in and out of metropolitan areas. To address this concern, it is useful to separate locations between metropolitan and non-metropolitan counties as defined by the Office of Management and Budget. The main criterion for a county to be classified as metropolitan is that it is part of an urban area of at least 50,000 residents.¹³ For 1980 we obtain that 90% of the counties with employment in manufacturing above 15214 (or 9.63 in logs) are metropolitan counties, whereas 90% of counties with employment in manufacturing below 4272 (or 8.36 in logs) are non-metropolitan counties. Comparing these numbers with the curve for 1980-2000 in Figure 7 makes clear that the pattern we document is not only a shift of manufacturing employment from cities to rural areas but a continuous dispersion throughout all county sizes. Similarly, we can compute the same thresholds for employment in services. For 1980 we find that 90% of counties above 22248 (or 10.01 in logs) employees in services are classified as metropolitan and 90% of counties with service employment levels below 16155 (9.69 in logs) are classified as non-metropolitan. Hence it is clear that the negative scale dependence at the top of the distribution in Figure 8 includes only counties that form integral part of cities. This is consistent with our argument that the negative scale dependence observed for large counties is the result of congestion costs in urban areas. In addition, the positive scale dependence in the middle part of the distribution includes both metro and non-metro counties.

In looking at Figure 8, one may argue that although the service sector as a whole exhibits the aforementioned S-shaped pattern in 1980-2000, particular industries within the service sector may not. The empirical evidence suggests that finance, insurance and real state (FIRE) is a sector where IT has been particularly important and so we expect to see the pattern there. Other important industries are retail and other services. Figure 9 presents the kernel regressions for these three industries. Clearly the S-shaped pattern is present in all of them, perhaps somewhat more pronounced in FIRE and other services. Unfortunately, service employment data is not available for the beginning of the twentieth century, so we cannot contrast the predictions of the model for services in this earlier period.¹⁴

[FIGURE 9 ABOUT HERE]

¹³To be precise, before 2003 metro areas were defined to include central counties with one or more cities of at least 50,000 residents or with an urbanized area of 50,000 or more and total area population of at least 100,000. Outlying counties were included if they were economically tied to the central counties. See <http://www.ers.usda.gov/Briefing/Rurality/NewDefinitions/> for more details.

¹⁴ All the same patterns for scale dependence in the growth rates are preserved if we focus on employment density (employment over county area). The same patterns are also present if we plot sectoral growth rates as a function of total, rather than sectoral, employment. This finding is important since in the theory some forces are sector specific (e.g., knowledge spillovers), whereas others are location specific and apply to all sectors (e.g., land rents).

The model above also has predictions for the evolution of the shares of services and manufacturing employment. Figure 10 presents the shares of employment in both the manufacturing and service industries in the U.S. in the past century. The data come from the Statistical Abstracts of the United States, published by the U.S. Bureau of the Census. The figure shows the well established increase in the service employment share. These graph should be compared to the results of our numerical simulation presented in the previous section and in particular to Figure 6 (keeping in mind that we did not include an agricultural sector in the theory).

[FIGURE 10 ABOUT HERE]

4.3 Manufacturing and Services in Europe

To make sure our findings are not specific to the United States, we also analyze spatial employment growth across European regions. Data on sectoral employment come from the Cambridge Econometrics Regional Database, which covers 236 Western European regions from 1975 to 2000.

Although we are unable to study the effect of electricity, our time series is long enough to analyze the effect of IT on different sectors of the economy. Before doing so, we need to compare the European experience to the U.S., both in terms of the sectoral impact and in terms of the timing.

Regarding the sectoral impact of IT, results are similar to those in the United States. Manufacturing is not benefitting to the same extent as services. As pointed out by Basu et al. (2003), although manufacturing accounts for about one fifth of GDP, it has less than one fifth of computers both in the United States and the United Kingdom. Within the broad category of services, those subsectors that have invested most heavily in IT in the U.K. are: finance/insurance, business services/real estate, and communications (Basu et al., 2003). Given that our database does not have disaggregated data on all of these sectors, in our empirical analysis we focus on ‘banking and insurance’.

Regarding the timing, there is ample evidence that Europe has lagged behind the U.S. in the adoption of IT (Gust and Marquez, 2002; van Ark et al., 2002). We therefore take 1985-2000 as the relevant period of IT diffusion in Europe.

In Figure 11 we pooled the data for 5-year intervals between 1985 and 2000 to increase sample size. As in the U.S., for Europe we observe a declining curve for manufacturing and an S-shaped curve for services (in this case the banking and insurance industry). This is exactly as the theory predicts for mature (manufacturing) and young (services) industries.

[FIGURE 11 ABOUT HERE]

4.4 Scale Dependence in Productivity Growth

In our theory the driving force behind the level and the growth of employment is the level and the growth of TFP. In our model TFP in a given industry can grow for two distinct reasons: knowledge spillovers from neighboring counties, and technology diffusion. If spillovers are weak or nonexistent and technology diffusion dominates, we get convergence across regions. Counties with initially lower levels of TFP experience faster TFP growth. If spillovers are strong enough though, the relation between TFP growth and initial TFP will reverse. This does not happen for counties at the bottom end of the TFP distribution. For those counties diffusion continues to dominate, and TFP growth exhibits negative scale dependence. Once we get to the middle of the TFP distribution, spillovers start to outweigh technology diffusion. Since spillovers become stronger with proximity to employment agglomerations in the sector, TFP will grow faster in counties with initially higher TFP. The result is divergence. This pattern will not continue all the way to the top part of the distribution though. Although spillovers continue to strengthen, they do so at a decreasing rate. This leads to a slowdown in TFP growth, so that convergence resurfaces.

In young industries, spillovers are strong, and we get the S-shaped curve described above: convergence at the bottom and at the top part of the distribution, with divergence in the middle. In mature industries, spillovers cease to dominate, while the technology is still not fully diffused. The non-monotonicity between TFP growth and initial TFP level (or, similarly, employment growth in the industry and initial employment level) disappears, and we obtain convergence across the entire distribution.

Focusing on the manufacturing sector, production per unit of land is given by $M(\ell) = Z_M(\ell)L_M(\ell)^\mu$, where $Z_M(\ell)$ is manufacturing TFP and $L_M(\ell)$ denotes labor per unit of land. Solving out for TFP, we get

$$Z_M(\ell) = \frac{M(\ell)}{L_M(\ell)^\mu}.$$

Similarly, in the case of services, TFP is given by

$$Z_S(\ell) = \frac{S(\ell)}{L_S(\ell)^\sigma}.$$

To compute sectoral TFP we need empirical counterparts of land, sectoral output, and sectoral employment. For land we use data on county area from the U.S. Geological Survey.¹⁵ Manufacturing output for the years 1900-1920 come from the Historical Census Browser at the University of Virginia, which provides county level data of the value of manufacturing production. Manufacturing and services

¹⁵This calculation of TFP does not account for capital or other factors of production, but has the advantage of being exactly consistent with our theory.

output for the years 1980-2000 comes from the Bureau of Economic Analysis, which collects county level data on total earnings per sector. Sectoral employment variables are the ones used in the previous section. Substituting these variables into the above equations allows us to compute TFP for manufacturing and services in each county. These measures of TFP depend on the value of the parameters μ and σ . As in the numerical simulations, we set $\mu = 0.5$ and $\sigma = 0.55$. The results are shown not to change qualitatively if we increase the values of μ and σ to, say, $2/3$.

[FIGURE 12 ABOUT HERE]

Figure 12 plots annual de-trended TFP growth in manufacturing on the log of initial TFP for the period 1900-1920. Given that manufacturing is young during this period, TFP growth exhibits the expected S-shaped pattern. By the last decades of the 20th century manufacturing has become a mature industry. Spillovers have lost their importance. Figure 13 shows how for the period 1980-2000 the S-shaped pattern has disappeared, with convergence now dominating across the entire distribution.

[FIGURE 13 ABOUT HERE]

If the S-shaped pattern is related to the youth of an industry, it should also apply to the service industry at the end of the 20th century. Figure 14 plots de-trended TFP growth for the period 1980-2000 for three different service sectors: retail, other services and FIRE. As expected, the S-shape re-emerges. Note one slight difference with the manufacturing industry during the period 1900-1920, the convergence at the top part of the distribution is less pronounced. This indicates that spillovers are sufficiently strong to cause high TFP growth in most of the highly productive counties. Of course, when we look at employment, land congestion costs for large agglomerations will yield more pronounced negative scale dependence at the top of the distribution for employment growth.

[FIGURE 14 ABOUT HERE]

Figures 12, 13 and 14 would not look qualitatively different if they were to plot TFP growth as a function of initial employment, rather than as a function of initial TFP. This is because employment and TFP are positively correlated in the data, as in our theory. In our minds, the fact that employment and TFP are positively related indicates that the driving force behind the patterns of employment and TFP that we have documented is the diffusion and adoption of technologies, as in the theory we propose.

4.5 Subsectoral Analysis

Our empirical analysis so far has been based on rather broad sectors. In principle, this makes sense for two reasons. First, at higher levels of disaggregation, one would expect the behavior of particular

sectors to become more idiosyncratic. Second, at higher levels of disaggregation, data availability at the county level becomes problematic because of disclosure and confidentiality issues. However, while taking into account these caveats, it may still be worthwhile to explore the relation between spatial growth and industry age for more detailed sectors.

The goal of this section, then, is to analyze whether our findings continue to hold for more disaggregated sectors. In particular, we look at the service sector (retail, FIRE and other services) and analyze whether in the last two decades those subsectors which were particularly affected by IT exhibited an S-shaped spatial growth pattern, and whether those subsectors which did not experience much effect from IT exhibited a monotonically decreasing spatial growth pattern.

We use employment data at the 2-digit SIC level from the County Business Patterns dataset spanning the time period 1977-1997.¹⁶ For the 60 available sectors, on average there are data for only about one third of the slightly more than 3000 counties in the United States. As mentioned, this is mainly a problem of confidentiality: when employment in a certain sector and county is concentrated in a limited number of firms, the data are not disclosed. To limit this problem, we focus on those subsectors for which we have at least two thirds of the counties. Within the three service categories (retail, FIRE, other services), this leaves us with ten subsectors.

To decide which sectors to focus on, we turn to the empirical literature on IT intensity. There are three relevant studies that analyze IT intensity at the 2-digit SIC level (Chun et al., 2005; Caselli and Paternò 2001; McGuckin and Stiroh, 2002). In all these studies IT intensity is defined as IT capital as a share of total capital.¹⁷ For each of those studies, we choose the most IT intensive sector and the least IT intensive sector, within the subset of sectors that have observations for at least two thirds of the counties. In spite of the differences in definitions, in all three studies the most IT intensive sector is ‘legal services’ and the least IT intensive sector is ‘auto repair’. For the case of ‘legal services’ the IT intensity is estimated to be around 30% in Chun et al. (2005) and Caselli et al. (2001) and 17% in McGuckin et al. (2002).¹⁸ For the case of ‘auto repair’ the IT intensity is found to be very low in all three studies, between 2% and 4%.¹⁹

[FIGURES 15 AND 16 ABOUT HERE]

¹⁶These data are available on-line at the Geospatial & Statistical Datacenter at the University of Virginia.

¹⁷The definitions of IT capital are slightly different across studies though: in Chun et al. (2005) and Caselli et al. (2001) it is essentially defined as the sum of hardware and software, whereas in McGuckin et al. (2002) it refers to the sum of computer hardware and other high-tech equipment.

¹⁸To be precise, Chun et al. (2005) finds a figure of 30% for the year 2000, Caselli et al. (2001) finds 29% for the year 1999, and McGuckin et al. (2002) reports 17% for 1996.

¹⁹Chun et al. (2005) reports 3.1% for the year 2000, Caselli et al. (2001) finds 2.3% for 1999, and McGuckin et al. (2002) gives a number of 3.8% for 1996.

Figure 15 and Figure 16 shows employment growth between 1977 and 1997 across U.S. counties, using the same methodology as before. As can be seen, ‘legal services’ exhibits the S-shaped spatial growth pattern. This is consistent with the importance of IT in that particular subsector of the economy. In contrast, ‘auto repair’ looks like a mature sector, with convergence across the entire distribution. Again, this is consistent with IT being of little importance in that particular sector. Although doing a more in-depth analysis at the 2-digit sector is beyond the scope of this paper, these examples do suggest that our basic finding — an S-shaped pattern of spatial growth in sectors affected by IT and a monotonically declining pattern of spatial growth in the rest of the economy — goes through when we look at more detailed sectors.

5 Conclusions

We have documented a new fact about the evolution of employment across sectors and industries. The spatial evolution across regions seems to be related to industry age. At a minimum, it is clear that the scale dependence in employment growth is different in the service and manufacturing sectors, and that the scale dependence in manufacturing at the turn of the twentieth century resembles the one in services in the last couple of decades of the twentieth century.

Our theory suggests that this distinct evolution in the manufacturing and service sectors may be related to the age of an industry as measured by the time since the last GPT innovation important for that sector. Young industries innovate and benefit from knowledge spillovers. This process leads, through trade, to changes in spatial specialization patterns, and technological diffusion, consistent with the observed scale dependence in employment growth in young industries. Old sectors disperse as technology disperses further and firms move to locations where land rents are low. Importantly, we have also documented in the data similar patterns of scale dependence for productivity. This is consistent with our theory where the driving force behind the observed employment patterns are spillovers and technological diffusion.

One caveat to our findings is that we did not document employment growth in the service industry in the first two decades of the twentieth century, a period where we would argue the service industry was old. Data limitations prevented us from doing so.

The theory presented endogenizes technological growth across regions by making it a function of the level of employment in nearby locations. However, the evolution of these technologies, and the technological spillovers themselves, are modeled only in reduced form. Modelling explicitly the adoption decisions of firms would, of course, lead to a richer spatial theory of endogenous growth.

References

- [1] Atkeson, A. and P.J. Kehoe, 2006. Modeling the Transition to a New Economy: Lessons from Two Technological Revolutions, Forthcoming in the *American Economic Review*.
- [2] Bartelsman, E.J. and Beaulieu, J.J., 2004. A Consistent Accounting of U.S. Productivity Growth, unpublished manuscript.
- [3] Basu, S. and Fernald, J., 2006. ICT as a General-Purpose Technology: Evidence from U.S. Industry Data, unpublished manuscript.
- [4] Basu, S., Fernald, J.G., Oulton, N., and Srinivasan, S., 2003. The Case of the Missing Productivity Growth: Or, Does information technology explain why productivity accelerated in the United States but not the United Kingdom?, Discussion Paper Number 2021, Harvard Institute of Economic Research.
- [5] Baumol, W.J., 1967. Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis, *American Economic Review*, 57, 415-426.
- [6] Bresnahan, T. F., and Trajtenberg, M., 1995. General Purpose Technologies: Engines of Growth?, *Journal of Econometrics*, 65, 83-108.
- [7] Bosworth, B.P. and Triplett, J.E., 2006. Services Productivity in the United States: Griliches' Services Volume Revisited, The Brookings Institution, unpublished manuscript.
- [8] Buera, F. and J. Kaboski, 2006, The Rise of the Service Economy, Northwestern University Working Paper.
- [9] Caselli, P. and Paternò, F., 2001. ICT Accumulation and Productivity Growth in the United States: An Analysis Based on Industry Data, Temi di Discussione 419, Banco d'Italia.
- [10] Chun, H., Kim, J.-W., Lee, J. and Morck, R., 2005. Information Technology, Creative Destruction, and Firm-Specific Volatility, unpublished manuscript.
- [11] David, P.A., 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox, *American Economic Review*, 80, 355-361.
- [12] David, P.A. and Wright, G., 2003. General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution, in: P.A. David and M. Thomas (eds.), *The Economic Future in Historical Perspective*, Oxford University Press.
- [13] Desmet, K. and M. Fafchamps, 2006, Employment Concentration across U.S. Counties, *Regional Science and Urban Economics*, 36, 482-509.
- [14] Duranton G. and D. Puga, 2001. Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products, *American Economic Review*, 91, 1454-1477.
- [15] Duranton, G., 2006, Urban evolutions: The fast, the slow, and the still, Forthcoming in the *American Economic Review*.
- [16] Forman, C, Goldfarb, A. and S. Greenstein, 2005. Geographic location and the diffusion of Internet technology, *Electronic Commerce Research and Applications*, 4, 1-13.
- [17] Forstall, R.L., 1996. *Population of the States and Counties of the United States: 1790 to 1990*, U.S. Bureau of the Census, Washington, D.C.
- [18] Forstall, R. L., 1996. *Population of States and Counties of the United States: 1790 to 1990*. Bureau of the Census. Washington, DC: U.S. Government Printing Office.
- [19] Gabaix, X., 1999, Zipf's Law for Cities: An Explanation, *Quarterly Journal of Economics*, 114, 739-67.

- [20] Griliches, Z., 1957. Hybrid Corn: An Exploration in the Economics of Technological Change, *Econometrica*, 25, 501-522.
- [21] Gust, C. and Marquez, J., 2002. International Comparisons of Productivity Growth: The Role of Information Technology and Regulatory Practices, International Finance Discussion Paper Number 727, Board of Governors of the Federal Reserve System.
- [22] Henderson, J. V., 2005, Urbanization and Growth, Chapter 24 in *Handbook of Economic Growth*, 1, Part B, 1543-1591, Elsevier.
- [23] Henderson, J. V. and A. Venables, 2005, The Dynamics of City Formation: Finance and Government, Brown University Working Paper.
- [24] Hobijn, B. and Jovanovic, B., 2001. The Information-Technology Revolution and the Stock Market: Evidence, *American Economic Review*, 91, 1203-1220.
- [25] Holmes, T., 2004. Step-by-step Migrations, *Review of Economic Dynamics*, 7, 52-68.
- [26] Jovanovic, B. and Rousseau, P.L., 2005. General Purpose Technologies, NBER Working Paper #11093.
- [27] Kendrick, J., 1961. *Productivity Trends in the United States*. Princeton: Princeton University Press.
- [28] Kongsamut, P., Rebelo, S. and X. Danyang, 2001. Beyond Balanced Growth, *Review of Economic Studies*, 68, 869-882.
- [29] Lucas R.E. Jr. and E. Rossi-Hansberg, 2002, On the Internal Structure of Cities, *Econometrica*, 70, 1445-1476.
- [30] Lucas, R. E. Jr., 2004, Life Earnings and Rural-Urban Migration, *Journal of Political Economy*, 112, S29-S59.
- [31] McGuckin, R.H. and Stiroh, K.J., 2002. Computers and Productivity: are Aggregation Effects Important?, *Economic Inquiry*, 40, 42-59.
- [32] Ngai, L.R. and Pissarides, c., 2004. Structural Change in a Multi-Sector Model of Growth, CEPR Discussion Paper No. 4763.
- [33] Rosenberg, N. and M. Trajtenberg, 2004. A General-Purpose Technology at Work: The Corliss Steam Engine in the Late- Nineteenth-Century United States, *Journal of Economic History*, 64, 1-39.
- [34] Rossi-Hansberg, E., 2004, Cities under Stress, *Journal of Monetary Economics*, 51, 903-937.
- [35] Rossi-Hansberg, E., 2005, A Spatial Theory of Trade, *American Economic Review*, 95, 1464-1491.
- [36] Rossi-Hansberg, E. and M. L. J. Wright, 2006, Urban Structure and Growth, Forthcoming in the *Review of Economic Studies*.
- [37] Triplett, J.E. and Bosworth, B.P., 2002. "Baumol's Disease" Has Been Cured: IT and Multifactor Productivity in U.S. Services Industries, The Brookings Institution, Washington, D.C.
- [38] van Ark, B., Melka, J., Mulder, N., Timmer, M., and Ypma, G., 2002. ICT Investment and Growth Accounts for the European Union, 1980-2000. *Final Report on ICT and Growth Accounting for the DG Economics and Finance of the European Commission*, Brussels.

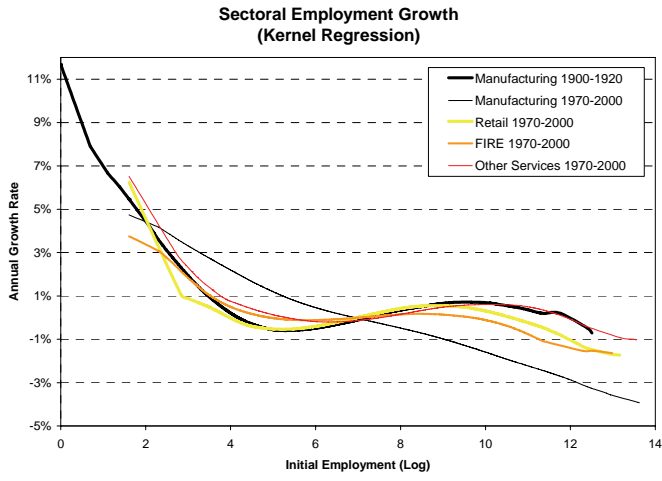


Figure 1

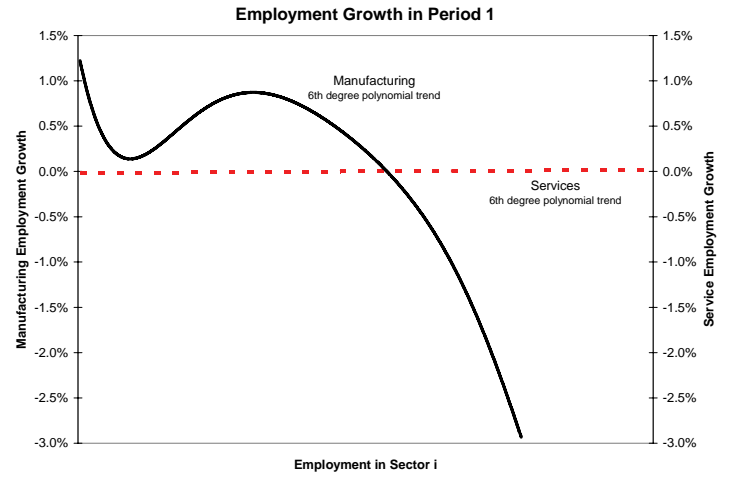


Figure 2a

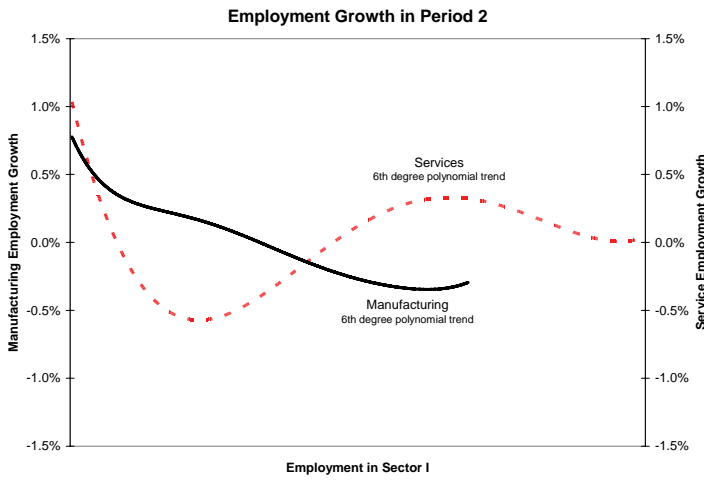


Figure 2b

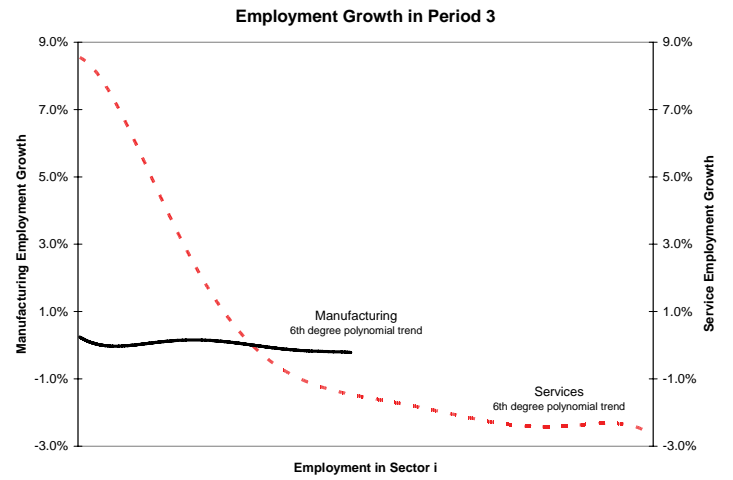


Figure 2c

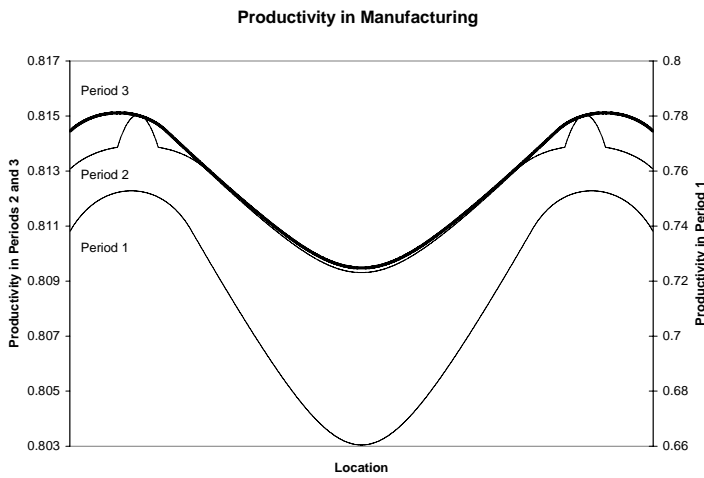


Figure 3

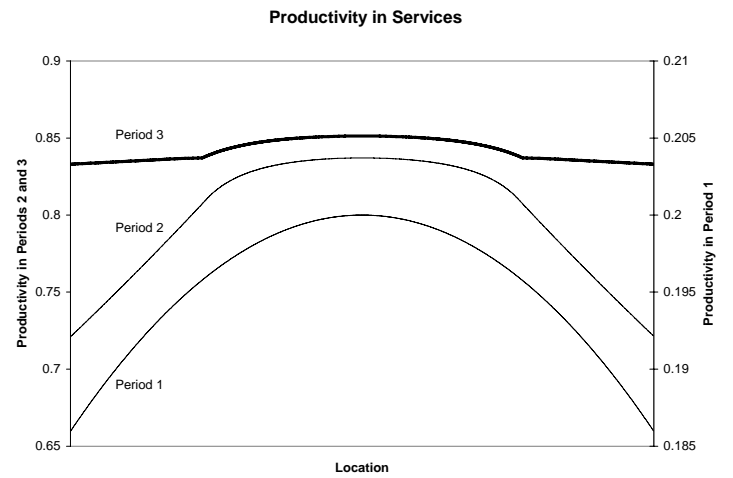


Figure 4



Figure 5

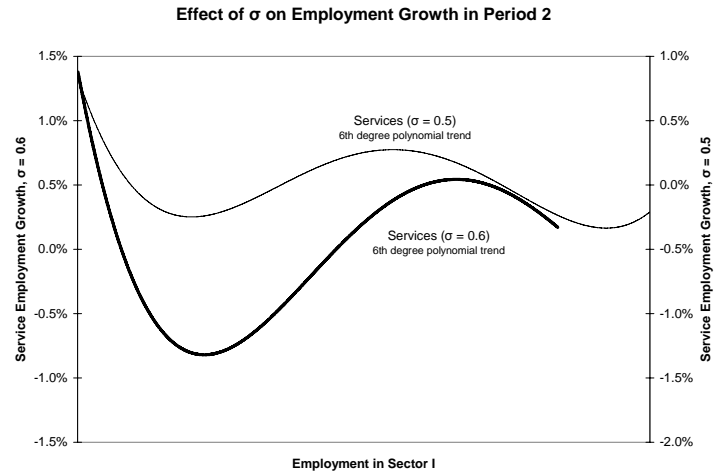


Figure 6

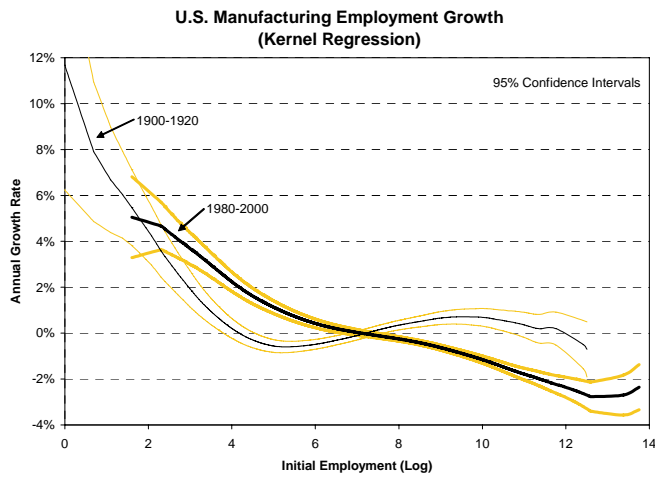


Figure 7

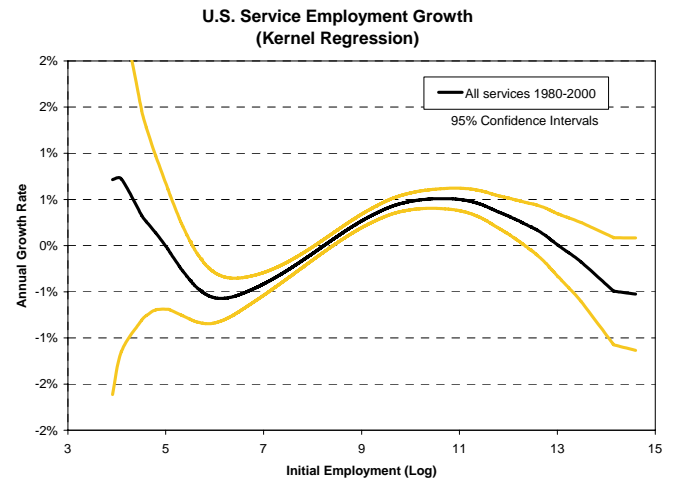


Figure 8

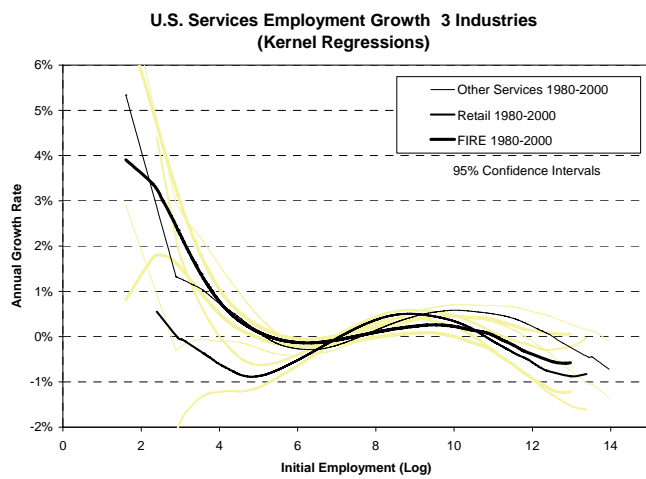


Figure 9

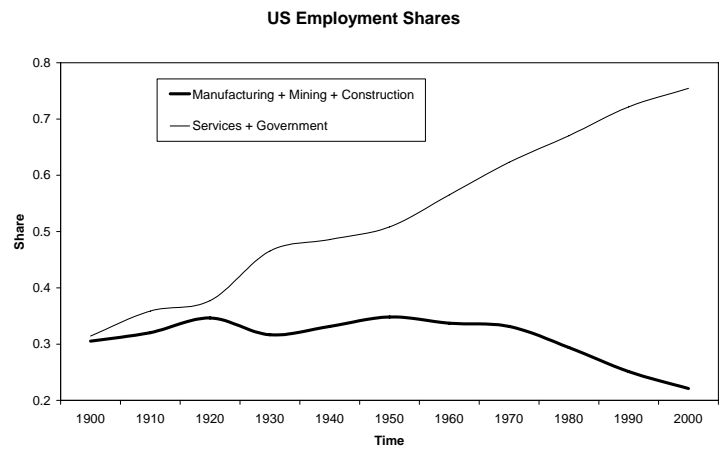


Figure 10

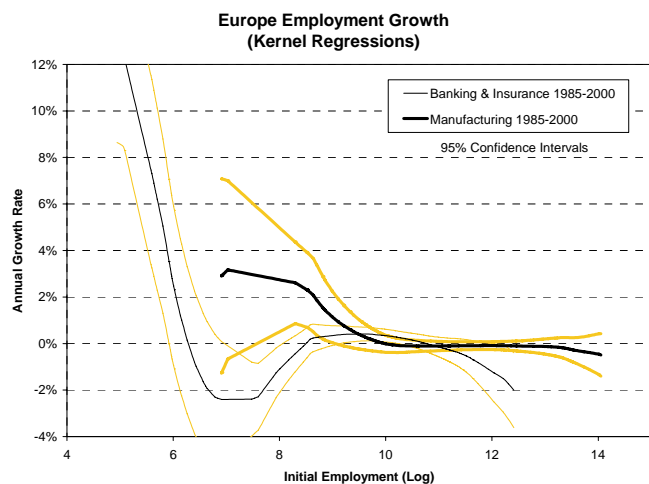


Figure 11

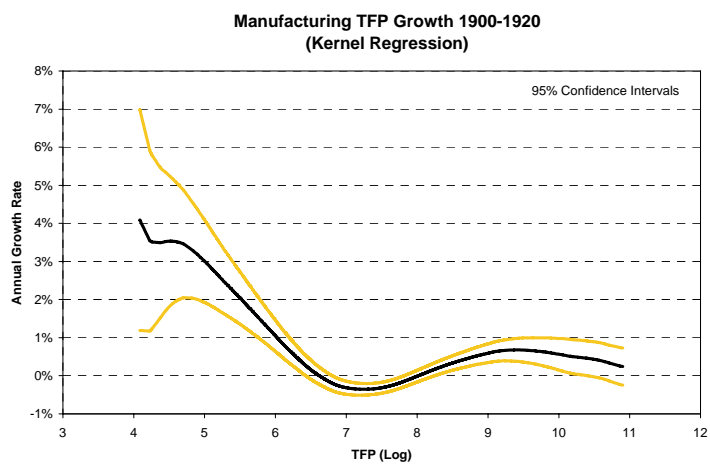


Figure 12

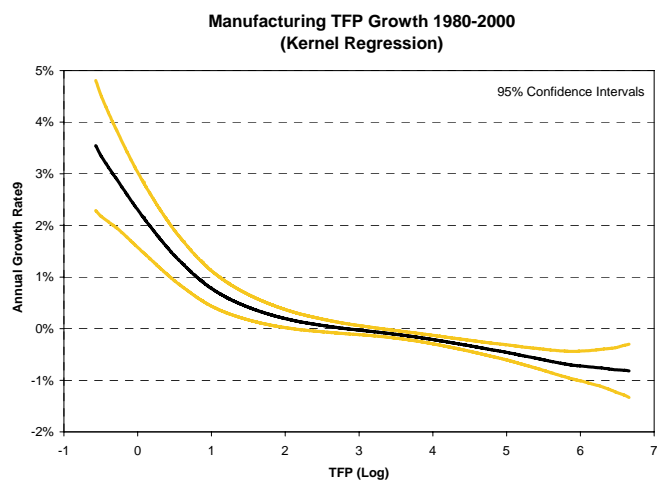


Figure 13

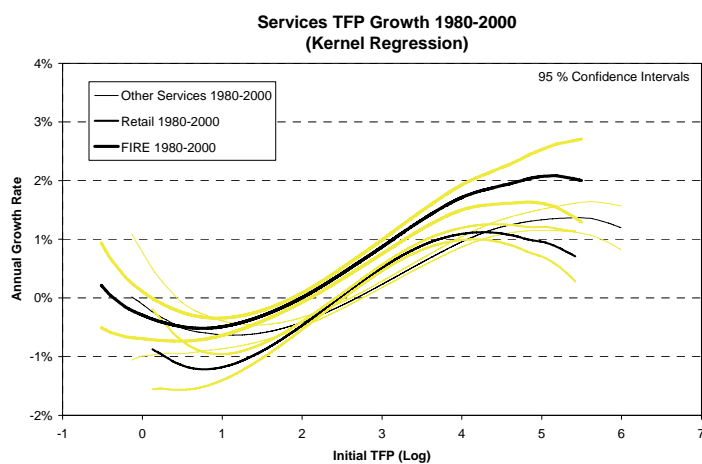


Figure 14

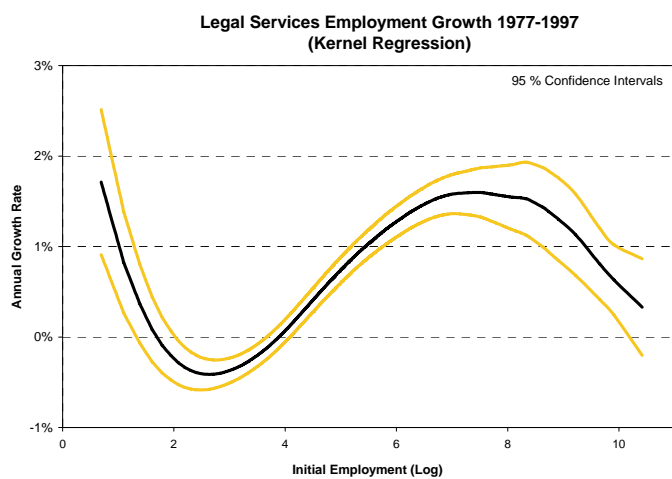


Figure 15

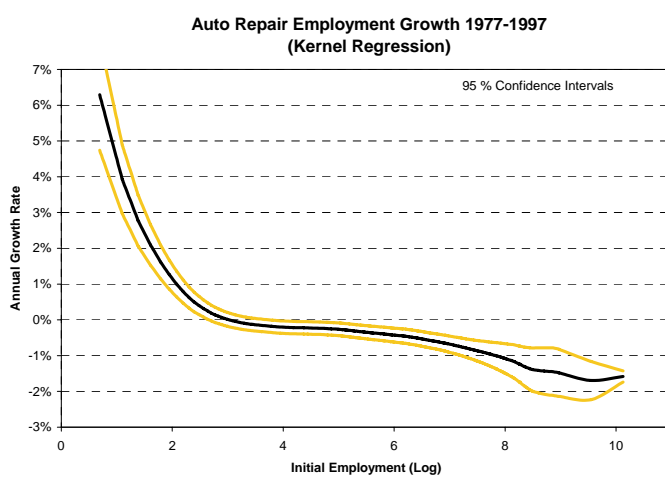


Figure 16